

# Three competing patterns

Rita Abraham and Jan Vrbik  
 Department of Mathematics  
 Brock University, 500 Glenridge Ave.  
 St. Catharines, Ontario, Canada, L2S 3A1

## Abstract

Assuming repeated independent sampling from a Bernoulli distribution with two possible outcomes  $S$  and  $F$ , there are formulas for computing the probability of one specific pattern of consecutive outcomes (such as  $SSFSS$ ) winning (i.e. being generated first) over another such pattern (e.g.  $SFSSFS$ ). In this article we will extend the theory to three competing patterns.

## 1 Completing a pattern

### 1.1 from scratch

Consider a sequence of *independent* Bernoulli trials, each having two potential outcomes,  $S$  (a success) with the probability of  $p$  and  $F$  (a failure) with the probability of  $q = 1 - p$ . Then, choose a specific pattern of  $m$  consecutive outcomes (such as  $SSFSS$ ), and define the PROBABILITY GENERATING FUNCTION (PGF for short) of the number of trials needed to generate this pattern *for the first time, from scratch* (let us call the corresponding random variable  $X$ ) as

$$F(s) \equiv \mathbb{E}(s^X) = \sum_{n=0}^{\infty} f_n s^n$$

where  $f_n$  is the probability of completing the pattern, for the first time, at Trial  $n$ .

If  $u_n$  is the probability of completing the same pattern (but not necessarily for the first time) at Trial  $n$ , we can relate the two sequences by

$$u_n = \sum_{i=0}^n f_i u_{n-i} \tag{1}$$

which is correct for any  $n \geq 1$  (but not for  $n = 0$ ) after setting  $u_0 = 1$  and  $f_0 = 0$  (to prove that, partition the sample space of the first  $n$  trials according to the trial - denoted  $i$  - at which the first occurrence of the pattern happened,

and use the total-probability formula). Here we are assuming (and this is quite crucial) that upon completing one occurrence of the pattern, we are *not allowed* to use any of its symbols to help build its next occurrence - we have to start 'from scratch'. Let  $U(s)$  be the sequence generating function (SGF) of the  $u_n$  probabilities, i.e.  $U(s) \equiv \sum_{n=0}^{\infty} u_n s^n$ ; note that  $U(1) = \infty$  - these probabilities do not constitute a distribution the way the  $f_n$  probabilities do.

Multiplying (1) by  $s^n$  and summing over  $n$  from 1 to infinity yields  $U(s) - 1$  (to account for the missing  $u_0$ ) on the LHS and  $U(s) \cdot F(s)$ , representing the CONVOLUTION of the two sequences, on the RHS (simplifying a convolution of two sequences in this manner is the main reason for using generating functions in this context). The resulting equation can be easily solved for  $F(s)$ , leading to

$$F(s) = \frac{U(s) - 1}{U(s)} \quad (2)$$

To utilize this formula, we must first find  $U(s)$ , which happens to be easier than finding, directly,  $F(s)$ . To achieve the former, we relate the probability of finding the *symbols* of the pattern (visualize SSFFS) at Trials  $n - m + 1$  to  $n$  (note that this does *not* necessarily mean that the corresponding pattern has been completed at Trial  $n$  - it may have been completed *earlier*, e.g. at Trial  $n - 4$  using our SSFFS example, which would prevent its completion at Trial  $n$ , due to the from-scratch requirement) to the probabilities of completing the corresponding pattern at *one* of these trials (in the case of SSFFS, only Trials  $n$  and  $n - 4$  are eligible), thus:

$$p^3 q^2 = u_n + u_{n-4} \cdot p^2 q^2 \quad (3)$$

which holds for any  $n \geq 5$  ( $n \geq m$  in general). Multiplying by  $s^n$  and adding over  $n$  from  $m$  to infinity (note that  $u_1$  to  $u_{m-1}$  must equal to 0), yields

$$\frac{p^3 q^2 s^5}{1 - s} = (1 + p^2 q^2 s^4) (U(s) - 1) \quad (4)$$

which can be easily solved for  $U(s) - 1$ , and then converted to  $F(s)$  using (2), getting

$$F(s) = \frac{1}{1 + (1 - s) \cdot \frac{1 + p^2 q^2 s^4}{p^3 q^2 s^5}}$$

Note that  $U(1)$  is infinite but  $F(1) = 1$ , as expected. The mean number of trials needed to generate the first occurrence of this pattern is obtained from

$$\mu \equiv F'(s=1) = \frac{1 + p^2 q^2}{p^3 q^2} \quad (5)$$

It is difficult to spell out the general form of (3) which would apply to any pattern, but the idea is (hopefully) quite clear. One should note that sometimes there is only the  $u_n$  term on the RHS of the equation (consider the SSFF pattern), sometimes we have all  $m$  terms (e.g. SSSS yields  $u_n + u_{n-1} \cdot$

$p + u_{n-2} \cdot p^2 + u_{n-3} \cdot p^3$ ). The best way to do this is to slide the pattern past itself to see how many perfect matches one gets (each one of these contributes exactly one term to the RHS).

## 1.2 with a head-start

To play two or more such patterns against each other (i.e. observing which of them happens first), we must also find the PGF of the *remaining* number of trials needed to generate the first occurrence of a pattern, given that the first few of its symbols are already there. Thus (using the old SSFFS example),  $F^{\text{SSF}}(s)$  is a PGF of the number of trials to generate SSFFS, assuming an SSF head start. This implies that, if we are lucky, we can complete SSFFS in only 2 more trials (i.e. now  $f_2^{\text{SSF}} = qp$ ), and that is pretty much the only help we can get from SSF in this case. Note that this may get more complicated in general - we may get another, ‘shorter’ help from the head-start string (consider FSSFF with the head start of FSSF - now we can finish the pattern in a single trial by getting an F, but if we get an S instead, we can still complete the pattern in only three more trials).

Using a similar approach to deriving (1), the corresponding formula now reads

$$u_n^{\text{SSF}} = \sum_{i=0}^n f_i^{\text{SSF}} u_{n-i} \quad (6)$$

valid for *all*  $n \geq 0$ , since now we take  $u_0^{\text{SSF}} = 0$  (this applies to all  $u_0$  with a superscript;  $u_0 = 1$  remains an exception). Multiplying (6) by  $s^n$  and summing over all  $n$  (this time, we *include*  $n = 0$ ) yields

$$F^{\text{SSF}}(s) = \frac{U^{\text{SSF}}(s)}{U(s)} \quad (7)$$

where  $U(s)$  was defined in the previous section. To get  $U^{\text{SSF}}(s)$ , we repeat the logic of (3), getting an identical

$$p^3 q^2 = u_n^{\text{SSF}} + u_{n-4}^{\text{SSF}} \cdot p^2 q^2 \quad (8)$$

for  $n \geq m$ . What changes is that, instead of the old  $u_0 = 1$ ,  $u_1 = u_2 = u_3 = u_4 = 0$ , we now have  $u_0^{\text{SSF}} = u_1^{\text{SSF}} = u_3^{\text{SSF}} = u_4^{\text{SSF}} = 0$  but  $u_2^{\text{SSF}} = pq$ . Multiplying (8) by  $s^n$  and summing over  $n$  from  $m$  to infinity yields

$$\frac{p^3 q^2 s^5}{1-s} = (U^{\text{SSF}}(s) - pq s^2) + p^2 q^2 s^4 \cdot U^{\text{SSF}}(s)$$

which can be easily solved for  $U^{\text{SSF}}(s)$  and consequently converted to

$$F^{\text{SSF}}(s) = \frac{(1-s+p^2 q s^3) p q s^2}{1-s+p^2 q^2 s^4 - p^2 q^3 s^5}$$

after some simplification. Note that, similarly to  $F(s)$ , the new  $F^{\text{SSF}}(s)$  also evaluates to 1 at  $s = 1$ . This is a universal property of these PGFs, indicating the any pattern will be generated with the probability of 1 sooner or later.

The corresponding mean number of flips is now slightly smaller than (5), namely

$$\mu^{\text{SSF}} = \left. \frac{d}{ds} F^{\text{SSF}}(s) \right|_{s=1} = \frac{1 + p^2 q^2 - pq}{p^3 q^2}$$

### 1.2.1 Another example

To find  $U^{\text{SSS}}(s)$  for the SSSS pattern we start with

$$p^5 = u_n^{\text{SSS}} + u_{n-1}^{\text{SSS}} \cdot p + u_{n-2}^{\text{SSS}} \cdot p^2 + u_{n-3}^{\text{SSS}} \cdot p^3$$

valid for  $n \geq 4$ , realize that  $u_0^{\text{SSS}} = u_2^{\text{SSS}} = u_3^{\text{SSS}} = 0$ ,  $u_1^{\text{SSS}} = p$ , and end up with (after multiplying the previous equation by  $s^n$  and summing over  $n$  from 4 to infinity):

$$\frac{p^4 s^4}{1-s} = (U^{\text{SSS}}(s) - ps) + ps \cdot (U^{\text{SSS}}(s) - sp) + p^2 s^2 \cdot (U^{\text{SSS}}(s) - sp) + p^3 s^3 \cdot U^{\text{SSS}}(s)$$

which can be easily solved for  $U^{\text{SSS}}(s)$  and converted to  $F^{\text{SSS}}(s)$ .

## 2 Playing 2 patterns against each other

### 2.1 from scratch

Let us consider two patterns which may not necessarily be of the same length, but neither of them is allowed to be a substring of the other.

We now modify our notation: let  $F_1(s)$  and  $F_2(s)$  will be the PGFs of the number of trials to generate Pattern 1 and Pattern 2 (respectively) for the first time *from scratch*, while  $F_{1|2}(s)$  assumes that Pattern 2 has just been completed and can be used as a head start to help generate Pattern 1; similarly, we define  $F_{2|1}(s)$ . The corresponding expected values will be denoted  $\mu_1$ ,  $\mu_2$ ,  $\mu_{1|2}$  and  $\mu_{2|1}$  respectively.

Thus, for example, if the first pattern is SSFFS and the second one is FSFSSF,  $F_1(s)$  and  $F_{1|2}(s)$  are the same as  $F(s)$  and  $F^{\text{SSF}}(s)$  of the previous section, since the first pattern can use only the last three symbols of the second pattern (namely SSF) as its head start. Similarly, the second pattern can use only the last two symbols of the first pattern, which then defines  $F_{2|1}(s)$ ; finding it, together with  $F_2(s)$ , is left as an exercise. Here we quote only the corresponding two means:

$$\begin{aligned} \mu_2 &= \frac{1 + p^3 q^2}{p^3 q^3} \\ \mu_{2|1} &= \frac{1 - p^2 q^3}{p^3 q^3} \end{aligned}$$

Let now  $X_{1\{2\}}(s)$  be the SGF of the probabilities that Pattern 1 wins over Pattern 2 at the completion of the  $n^{\text{th}}$  trial, and let  $X_{2\{1\}}(s)$  be its vice-versa counterpart. Clearly,  $x_{1\{2\},0} = x_{2\{1\},0} = 0$ .

The  $f_{1,n}$  probability (of Pattern 1 completed, for the first time, at Trial  $n$ ) can be expanded as follows:

$$f_{1,n} = x_{1\{2\},n} + \sum_{i=0}^n x_{2\{1\},i} \cdot f_{1|2,n-i} \quad (9)$$

which is now correct for *all*  $n \geq 0$ . The RHS applies the total probability formula to the sample space of the first  $n$  trials, partitioning it according to the trial (denoted  $i$ ) at which Pattern 2 wins the game, adding the probability (the first term of the RHS) that Pattern 2 has not been completed yet, in which case Pattern 1 has won, at Trial  $n$ . Multiplying the previous equation by  $s^n$  and summing over  $n$  (from 0 to infinity) yields

$$F_1(s) = X_{1\{2\}}(s) + X_{2\{1\}}(s) \cdot F_{1|2}(s)$$

since the last term of (9) is a *convolution* of the two sequences, becoming a *product* of the corresponding generating functions, as explained earlier.

The same must be true with Patterns 1 and 2 interchanged, thus:

$$F_2(s) = X_{2\{1\}}(s) + X_{1\{2\}}(s) \cdot F_{2|1}(s)$$

The last two equations are easily solved for

$$\begin{aligned} X_{1\{2\}}(s) &= \frac{F_1(s) - F_2(s) \cdot F_{1|2}(s)}{1 - F_{1|2}(s) \cdot F_{2|1}(s)} \\ X_{2\{1\}}(s) &= \frac{F_2(s) - F_1(s) \cdot F_{2|1}(s)}{1 - F_{1|2}(s) \cdot F_{2|1}(s)} \end{aligned} \quad (10)$$

The probability that Pattern 1 wins (at some trial) is given by  $X_{1\{2\}}(1)$ , or more accurately (since a simple evaluation would lead to an indefinite answer of  $\frac{0}{0}$ ) by

$$\lim_{s \rightarrow 1} X_{1\{2\}}(s) = \frac{\mu_2 - \mu_1 + \mu_{1|2}}{\mu_{1|2} + \mu_{2|1}}$$

Applied to our example of playing SSFFS against FSFSSF this yields

$$\frac{1 - pq^3(1 + p)}{1 + q^2 + p^2q}$$

which increases from 50% in the  $p \rightarrow 0$  limit to 100% when  $p \rightarrow 1$  (note that Pattern 1 consists of the same number of Ss as Pattern 2, but has fewer Fs).

## 2.2 with a head start

To get ready for playing *three* patterns against each other (the main topic of this article), we have to extend the previous two formulas by assuming that the game (of playing Pattern 1 against Pattern 2) started from a completed Pattern 3 (and allowing either of the two competing patterns to utilize any of its symbols). By the same reasoning, it is easy to find that

$$\begin{aligned} X_{1\{2\}|3}(s) &= \frac{F_{1|3}(s) - F_{2|3}(s) \cdot F_{1|2}(s)}{1 - F_{1|2}(s) \cdot F_{2|1}(s)} \\ X_{2\{1\}|3}(s) &= \frac{F_{2|3}(s) - F_{1|3}(s) \cdot F_{2|1}(s)}{1 - F_{1|2}(s) \cdot F_{2|1}(s)} \end{aligned} \quad (11)$$

where  $X_{1\{2\}|3}(s)$  is the SGF of the probabilities of Pattern 1 winning over Pattern 2 in exactly  $n$  trials, *given* that the game has started from a completed Pattern 3.

## 3 Playing 3 patterns

We can extend (9) by expanding the probability of Pattern 1 beating Pattern 2 at a completion of Trial  $n$  (the LHS), partitioning the sample space according to the trial (denoted  $i$ ) at which Pattern 3 has been completed *for the first time*, including the possibility (the first term on the RHS), that Pattern 3 has not been completed yet. In a similar manner to deriving (9), this leads to

$$X_{1\{2\}}(s) = X_{1\{2,3\}}(s) + X_{3\{2,1\}}(s) \cdot X_{1\{2\}|3}(s) \quad (12)$$

where  $X_{1\{2,3\}}(s)$  is the SGF of the probability of Pattern 1 beating both Patterns 2 and 3 at the completion of the  $n^{\text{th}}$  trial.

Reversing the rôle of Patterns 1 and 3 we get

$$X_{3\{2\}}(s) = X_{3\{2,1\}}(s) + X_{1\{2,3\}}(s) \cdot X_{3\{2\}|1}(s)$$

Solving the last two equations, one gets

$$\begin{aligned} X_{1\{2,3\}}(s) &= \frac{X_{1\{2\}}(s) - X_{3\{2\}}(s) \cdot X_{1\{2\}|3}(s)}{1 - X_{3\{2\}|1}(s) \cdot X_{1\{2\}|3}(s)} \\ X_{3\{2,1\}}(s) &= \frac{X_{3\{2\}}(s) - X_{1\{2\}}(s) \cdot X_{3\{2\}|1}(s)}{1 - X_{1\{2\}|3}(s) \cdot X_{3\{2\}|1}(s)} \end{aligned}$$

and, by permuting the indices, an analogous solution for  $X_{1\{3,2\}}$ ,  $X_{3\{1,2\}}$ ,  $X_{2\{1,3\}}$  and  $X_{2\{3,1\}}$  (at this point it would appear that  $X_{1\{2,3\}}(s)$  is different from  $X_{1\{3,2\}}(s)$ , but keep on reading).

Utilizing (10) and (11), we can then express any of these solutions directly in terms of the  $F_i(s)$  and  $F_{i|j}(s)$  functions, getting (for expedience, we quote

each  $F$  without its  $(s)$  argument):

$$X_{1\{2,3\}}(s) = \frac{F_1(1 - F_{2|3}F_{3|2}) - F_2(F_{1|2} - F_{1|3}F_{3|2}) - F_3(F_{1|3} - F_{1|2}F_{2|3})}{1 - F_{1|2}F_{2|1} - F_{1|3}F_{3|1} - F_{2|3}F_{3|2} + F_{1|2}F_{2|3}F_{3|1} + F_{1|3}F_{3|2}F_{2|1}} \quad (13)$$

and its index-permuted equivalents. Now, it becomes explicitly obvious that  $X_{1\{2,3\}}(s) = X_{1\{3,2\}}(s)$  as expected (winning over Patterns 2 and 3 is the same as winning over Patterns 3 and 2).

To get the probability of Pattern 1 winning over the other two patterns at *any* trial, one has to evaluate (13) at  $s = 1$ . This yields (with the help of L'Hospital rule, having to differentiate each numerator and denominator *twice*)

$$\lim_{s \rightarrow 1} X_{1\{2,3\}}(s) = \frac{\mu_1(\mu_{2|3} + \mu_{3|2}) + \mu_2(\mu_{1|2} - \mu_{1|3} - \mu_{3|2}) + \mu_3(\mu_{1|3} - \mu_{1|2} - \mu_{2|3}) + \mu_{2|3}\mu_{3|2} - \mu_{1|3}\mu_{3|2} - \mu_{1|2}\mu_{2|3}}{\mu_{1|2}\mu_{2|1} + \mu_{1|3}\mu_{3|1} + \mu_{2|3}\mu_{3|2} - \mu_{1|2}\mu_{2|3} - \mu_{1|3}\mu_{3|2} - \mu_{2|1}\mu_{1|3} - \mu_{2|3}\mu_{3|1} - \mu_{3|1}\mu_{1|2} - \mu_{3|2}\mu_{2|1}} \quad (14)$$

One can get the other two answers by permuting indices.

Finally, the PGF of the game's duration is clearly given by the following sum

$$H(s) \equiv X_{(1|2,3)}(s) + X_{(2|1,3)}(s) + X_{(3|1,2)}(s)$$

The expected duration of the game is thus equal to the  $s$  derivative of this expression, evaluated (with the help of L'Hospital rule and the *fourth* derivative of the numerator and denominator) at  $s = 1$ . This yields

$$\lim_{s \rightarrow 1} H'(s) = \frac{\mu_1(\mu_{2|3}\mu_{3|2} - \mu_{2|3}\mu_{3|1} - \mu_{3|2}\mu_{2|1}) + \mu_2(\mu_{1|3}\mu_{3|1} - \mu_{1|3}\mu_{3|2} - \mu_{3|1}\mu_{1|2}) + \mu_3(\mu_{1|2}\mu_{2|1} - \mu_{1|2}\mu_{2|3} - \mu_{2|1}\mu_{1|3}) + \mu_{1|2}\mu_{2|3}\mu_{3|1} + \mu_{1|3}\mu_{3|2}\mu_{2|1}}{\mu_{1|2}\mu_{2|1} + \mu_{1|3}\mu_{3|1} + \mu_{2|3}\mu_{3|2} - \mu_{1|2}\mu_{2|3} - \mu_{1|3}\mu_{3|2} - \mu_{2|1}\mu_{1|3} - \mu_{2|3}\mu_{3|1} - \mu_{3|1}\mu_{1|2} - \mu_{3|2}\mu_{2|1}} \quad (15)$$

### 3.1 Example

Adding FSSSF to the two patterns of the previous section, we compute

$$\begin{aligned} \mu_3 &= \frac{1 + p^3q}{p^3q^2} \\ \mu_{3|1} &= \mu_3^{\text{FS}} = \frac{1 - p^2q^2}{p^3q^2} \\ \mu_{3|2} &= \mu_3^{\text{F}} = \frac{1}{p^3q^2} \\ \mu_{1|3} &= \mu_1^{\text{SSF}} = \mu_{1|2} \\ \mu_{2|3} &= \mu_2^{\text{F}} = \frac{1}{p^3q^3} \end{aligned}$$

Based on (14), the probability of SSFFS winning the game is

$$\frac{1 - pq^2(1+p)(1+q)}{3q + p^2(2+q)}$$

which varies from  $\frac{1}{3}$  at  $p \rightarrow 0$  (all three patterns have the same chance of winning, since they contain the same number of Ss, and that is all what counts in this limit) to its smallest value of 28.59% at  $p = 0.2495$ , to its largest value of 50% at  $p \rightarrow 1$  (Patterns 1 and 3 have the same chance of winning, each containing two Fs; Pattern 2 with three Fs is now out of the contest).

The expected duration of the game is

$$\frac{1 + p^2q(1 - pq^3(1 + pq))}{p^3q^2(3q + p^2(2 + q))}$$

which reaches its smallest value of 15.88 trials at  $p = 0.5796$ ; in each of the  $p \rightarrow 0$  and  $p \rightarrow 1$  limits, this mean (or average) number of trials to complete this game becomes infinite.

### 3.2 Final challenge

It would be an interesting (but certainly non-trivial) exercise to extend the theory to four or more competing patterns.

## References

- [1] Feller W: *An Introduction to Probability Theory and Its Applications*, Volume I, Third Edition, John Wiley & Sons, 1968
- [2] Vrbik J: "Betting two patterns against each other", *The Mathematica Journal*, **13**, 2011